CUSTOMER SEGMENTATION USING DATA SCIENCE

Stage 1 : Data collection

Collect customer data:

- Gather customer data from diverse sources such as customer databases , transaction , social media and surveys to enhance segmentation accuracy
- Determine the key variables for effective customer segmentation , considering factors like demographics , purchase history , and online behavior

Data preprocessing and cleaning

- Cleanse and preprocess data to ensure accuracy ,completeness and consistency and to remove errors , duplicates and missing values
- Integrate data from CRM systems , online platforms and other sources to create a comprehensive customer profile

Consent and privacy compliance

- Adhere to data privacy regulations and obtain explicit customer consent before collecting and using their data
- Implement robust security measures , including encryption and access controls to protect customer data

Stage 2 : Data preprocessing

Label encoding

- Utilize label encoding to convert categorical features (e.g., product categories) into numerical representations.
- Ensure that each unique category is assigned a distinct integer label , facilitating the preparation of data for clustering or classification algorithms.

Transformation of data

- Employ scaling or normalization techniques on numerical features to ensure uniform impact on the analysis.

- Consider method such as standardization (subtracting mean and dividing by standard deviation) or min-max scaling (scaling values between 0 and 1)

Splitting of data

- Execute a split on the dataset , creating distinct training and testing sets for model evaluation.
- Common split ratios include 70-30 , 80-20 , or 90-10 for training-testing divisions.
- Employ k-fold cross-validation where the dataset is divided into k subsets and the model undergoes training and testing k times.

Stage 3 : Feature engineering

Feature selection:

- Conduct a comprehensive analysis to pinpoint attributes and features that significantly contribute to the accuracy of customer segmentation . Employ statistical tests , correlation analysis , or domain expertise to discern relevant features

Feature creation

- Aggregate the total spending for each customer based on their transaction history
- Calculate the frequency of customer purchases over a defined period
- Determine the time elapsed since a customer's last purchase.
- Compute the average transaction amount for each customer.
- Augment the segmentation granularity by introducing features that offer a more enhanced understanding of customer behavior.

Encoding categorical variables

- Employ encoding methods like one-hot encoding to represent categorical variables as binary vectors.
- Ensure all categorical features are appropriately encoded to facilitate compatibility with machine learning algorithms
- Validate that all features, including newly created ones, are in a format suitable for machine learning model input.

Stage 4 : Cluster the data (K-Means clustering algorithm)

Specify number of clusters(K):

- Determine the optimal number of clusters(K) based on domain-knowledge or data-driven methods like silhouette score or elbow method

Initialize centroids

- Shuffle the dataset and randomly select K data points without replacement to serve as initial centroids.This establishes starting points for the iterative k-means clustering process

Iterative clustering

- Measure distances between data points and centroids, assigning each point to the cluster with the nearest centroid
- Compute new centroids as the mean of all data points assigned to each cluster.
- Iterate assignment and centroid recalculation until convergence, signaled by stable centroids.

Stage 5 : Visualization

- Create scatter plots to showcase how clusters are distributed concerning two chosen features. This provides insights into the relationships between clusters and specific attributes.
- Utilize bar charts to represent the sizes or distributions of different customer segments. This provides a clear visual understanding of the composition of each cluster
- Heatmaps to show similarities between segments

Stage 6 : Interpretation

Characteristics analysis

- Conduct a detailed analysis of each customer segment to gain insights into their unique characteristics, behaviors, and needs.
- Explore specific features that differentiate one segment from another.
- Consider demographic information, transactional behavior, and any other relevant factors contributing to the distinctiveness of each segment.

Deriving Insights

- Identify key factors that distinguish each customer segment from others. Uncover insights into what makes each segment unique and how they differ in terms of preferences, behaviors, or requirements.
- Utilize the gained insights to customize marketing strategies, refine product offerings, and enhance overall customer experiences.